Available online at www.sciencedirect.com

SCIENCE DIRECT®

Vision Research 45 (2005) 2287–2297

Vision Research

www.elsevier.com/locate/visres

Configural masking of faces: Evidence for high-level interactions in face perception

Gunter Loffler [a,*], Gael E. Gordon [a], Frances Wilkinson [b], Deborah Goren [b], Hugh R. Wilson [b]

[a] *Department of Vision Sciences, Glasgow Caledonian University, Cowcaddens Road, Glasgow G4 0BA, Scotland, UK*
[b] *Centre for Vision Research, York University, 4700 Keele Street, Toronto, ON, Canada M3J 1P3*

Received 20 May 2004; received in revised form 9 February 2005

Abstract

The perception of a stimulus can be impaired when presented in the context of a masking pattern. To determine the timing and the nature of face processing, the effect of various masks on the discriminability of faces was investigated. Results reveal a strong configural effect: the magnitude of masking depends on the similarity between mask and target. Masking is absent for non-face masks (noise, houses), modest for scrambled and inverted faces and strongest for upright faces, even when they differ in size, gender or viewpoint from the targets. This suggests an extra-striate location for the masking (possibly FFA). Reduced but significant masking for isolated face parts (internal features or head shape) is consistent with holistic computations in face perception. The duration over which a face mask can impair face discrimination (130 ms) is markedly longer than previously assumed and is sufficient for iterative and feedback computations to be part of face processing.

Keywords: Face perception; Masking; Temporal dynamics; Psychophysics

1. Introduction

Faces are extraordinarily complex and highly important visual stimuli. Social interactions depend critically on their correct recognition and interpretation. Perhaps unsurprisingly, primates and humans have evolved specialised processing areas: the inferior temporal cortex (IT) and the superior temporal sulcus (STS) in monkeys (Desimone, 1991; Gross, 1992; Gross, Rocha-Miranda, & Bener, 1972) and the fusiform face area (FFA) in humans (Allison, Puce, Spencer, & McCarthy, 1999; Kanwisher, McDermott, & Chun, 1997; Sergent & Signoret, 1992).

Despite extensive research, the exact nature of the computations underlying our remarkable ability to discriminate faces is still unclear. Insight into these computations can be gained by studying the duration of face processing. The precise timing is an important parameter as it can be used to distinguish between computational strategies. Rapid processing has been taken as evidence for purely feed-forward computations, too fast for feedback to be involved (Lehky, 2000); prolonged durations allow for iterative and recurrent processing (Lamme & Roelfsema, 2000). This has strong implications for the neural hardware. The neuronal implementation of a purely feed-forward computation would have to be largely hard-wired. This is not a requisite for iterative processing.

Of equal importance in elucidating face perception is how facial information is combined. Faces appear to be computed differently from other objects (Diamond & Carey, 1986; Tanaka & Farah, 1993). It is well documented

* Corresponding author. Tel.: +44 141 331 3386.
E-mail address: gloe@gcal.ac.uk (G. Loffler).

0042-6989/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2005.02.009

that in normal circumstances human observers perceive faces holistically, relying on processing the relations among facial features correctly placed within a head shape (Diamond & Carey, 1986). During learning and subsequent recognition of intact upright faces, observers do not appear to emphasize explicit representation of face parts, which contrasts with recognition for houses, scrambled or inverted faces (Tanaka & Farah, 1993). It has been proposed that faces are represented as undifferentiated whole shapes, with little or no explicit representation of face parts. However, humans can also recognise a face on the basis of isolated features presented independently of the facial context or within a different context (e.g. scrambled faces), albeit with some loss of accuracy (Tanaka & Farah, 1993). It appears then that both feature based and holistic representations can be used in face discrimination and their dependence or independence has been a matter of debate (Collishaw & Hole, 2000; Tanaka & Sengco, 1997).

Both questions can be addressed with the psychophysical tool of visual masking. Masking is the phenomenon in which the sensitivity to a test stimulus is impaired by a second, masking stimulus (Breitmeyer, 1984). Masking may occur when target and mask are presented simultaneously but also when the mask follows the target presentation (backward masking). Usually, the gap between target and mask has to be short for masking to occur, and the effect is often restricted to 40 ms or less as in the case of pattern masking (e.g. Kovacs, Vogels, & Orban, 1995). Masking effects have been explained under the assumption that a mask following a stimulus creates a transient. If the target computation is incomplete when the mask is presented, it can interrupt the processing of the target and thus impair perception. Consequently, the temporal window over which a mask can impair perception is thought to reflect the duration of the underlying cortical computation.

Masking has enjoyed tremendous success in pattern vision (Regan, 2000) but limited use has been made of masking methods with face stimuli. The exceptions (Costen, Shepherd, Ellis, & Craw, 1994; Esteves & Ohman, 1993; Moscovitch & Radzins, 1987) have provided interesting insights into face processing. For example, the existence and the site of a central face processor was localised through masking in the right hemisphere, opposite to the left preference for word recognition (Moscovitch & Radzins, 1987). This observation preceded more recent brain imaging studies, which confirmed the right hemispheric dominance in face perception for right-handed subjects (Kanwisher et al., 1997; Sergent & Signoret, 1992).

The duration necessary to analyze faces perceptually has been disputed. On one hand, subjects require at least 100–150 ms to correctly identify the emotion of faces (Esteves & Ohman, 1993) or to identify famous individuals (Costen et al., 1994). On the other hand, the discrimination of morphed face photographs is possible for presentations as short as 50–100 ms (Lehky, 2000). This could be because some aspects of face processing (e.g. discrimination of faces) are completed rapidly while others (e.g. recognition of emotions and identification of individual faces) are not. Our results provide an alternative solution to this issue by showing that the duration of masking depends on the type of mask used.

Various models have been proposed for the detrimental effect of backward masking. These include visual integration of events occurring in close spatio-temporal proximity (Di Lollo, 1980; Turvey, 1973), interruption of processing (Turvey, 1973) and competitive neural interactions (Breitmeyer & Ganz, 1976). Such theories have been used to explain results on metacontrast masking and attributed to local contour interactions (e.g. Enns, 2004) that presumably occur at an early visual stage. None of these theories captures the effect, reported recently in a letter identification experiment, that the similarity between target and mask can influence the magnitude of masking (Enns, 2004): letters exerted a stronger masking effect than digits and noise, with isolated dots having no masking effect.

We aimed to explore the possibility that such a similarity-dependent masking effect may also be present for face perception. If it existed, this effect could be used to probe the computations underlying face processing. Face stimuli are particularly well suited to this approach since it is possible to present masks that are physically similar to the target (scrambled faces with the position of the internal features randomised) but perceptually different. Our results show that the amount of masking does indeed depend on the degree of similarity between face target and mask. Object similarity (face-ness) is required for strong masking to occur.

The stimuli in this study were simplified face images. Most previous studies of face perception employed photographs, computer averages of several photographs, or reconstructions from laser scanned faces (see Bruce & Young, 1998). Relating perception to the responses of underlying neural mechanisms is problematic with such highly complex stimuli. We have recently developed simplified face stimuli to reduce the complexity inherent in face photographs (Wilson, Loffler, & Wilkinson, 2002). The stimuli were designed to capture the major geometric aspects of faces (head shape, hair line, internal features size and placement) extracted from individual human face photographs (Fig. 1a) while omitting cues such as hair and skin texture, skin colour, wrinkles, etc. The rationale is that this geometric information is all that is available at distances greater than about 10 m, a distance at which face recognition is still easily performed. Any information contained in high spatial frequencies (e.g. texture), which have been shown to play a minor role in face recognition (Nasanen, 1999),

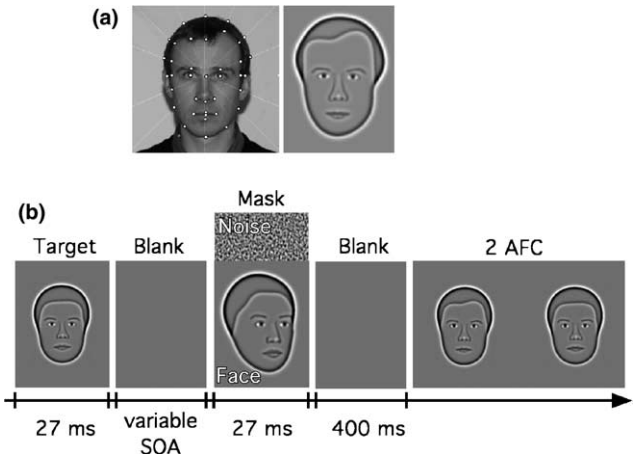

Fig. 1. Face stimuli and procedure. (a) Synthetic faces were created by extracting the major geometric information (head shape, hairline, shape and placement of internal features) from grey-scale photographs and reconstructing faces that were subsequently band-pass filtered (see text for details). (b) Face discrimination was measured using these synthetic faces in the presence of masking stimuli. Target faces were presented briefly (27 ms) following, or followed by, a mask (e.g. face or noise) for the same short duration. The face mask was always bigger in size than the target, shown from a side-view and of opposite gender. The onset time of the mask relative to the target (SOA) was varied. Subjects indicated which of two subsequently shown faces matched the target (2AFC).

is lost. We have shown that these faces combine simplicity and low-dimensional description with sufficient realism to permit individual identification (Wilson et al., 2002). Moreover, a comparable fMRI signal in the FFA shows that the brain processes these synthetic faces in a similar way to real face photographs (Loffler, Wilkinson, Yourganov, & Wilson, 2004).

2. Methods

2.1. Synthetic faces

The design of synthetic face stimuli has been described in detail elsewhere (Wilson et al., 2002). Briefly, the major geometric information is digitized at specific points from individual face photographs with neutral expressions (Fig. 1a). All face coordinates were measured relative to the bridge of the nose, which served as centre of a polar coordinate system. The radial coordinate was used to sample the head shape at 16 equally spaced points. A curve, consisting of a sum of seven radial frequencies (Wilson et al., 2002), was fitted to these points. The hairline was synthesized in the same way by nine points above the midline. A further 14 measurements defined the facial features (e.g. position of the eyes, length of the nose, width of the mouth). The position of all features was idiosyncratic. The shape of some features (eyes, eyebrows) was generic and identical in all faces while others (mouth, nose) used generic forms that were altered in width and length depending on the individual measurements. In total, each face was completely defined by 37 parameters and represented by a 37-dimensional vector. To avoid the physical size of a face serving as a potential cue for face discrimination, all faces were scaled to equal size. This was achieved by normalising all face measurements by the ratio between individual head radius and mean radius of the gender to which the face belonged. The images were subsequently band-pass filtered at the optimal spatial frequency for face identification (10 cycles/face width, Gold, Bennett, & Sekuler, 1999; Nasanen, 1999), with an optimal bandwidth of 2.0 octaves (circular DOG filter, Nasanen, 1999). The resulting faces (Fig. 1a) accentuate geometric information in the most important frequency band, while omitting such face cues as hair and skin texture, skin colour, wrinkles, etc.

A new set of 25 synthetic faces was presented in each experimental run. These face-sets were created by selecting one face of one gender randomly from the database of 80 faces. The geometric difference between the mean face, which always served as origin, and this face was then calculated. Here and elsewhere, differences between any two synthetic faces were mathematically defined as the Euclidean distance between the two 37-dimensional vectors representing each face. We have previously shown that a Euclidean norm represents perceptual differences between synthetic faces accurately (Wilson et al., 2002). The vector representing the selected face was then normalised to give the desired geometric difference from the mean, appropriate for the task and subject. Four equidistant vectors were placed between the mean and the selected face, to give a total of 5 faces along the dimension given by the selected face. For descriptive purposes, this dimension corresponds to an identity axis according to the face space proposed by Valentine (1991): faces change their distinctiveness but not their identity along such a dimension. Next, three additional faces were randomly selected from the same gender. Their vectors were first made orthogonal to each other (employing the Gram–Schmidt procedure) and the resulting axes also normalised and subdivided into equidistant vectors. A fifth axis was calculated as the principle diagonal of this 4D sub-space of our 37-dimensional face space. In each experimental trial, 2 faces were randomly selected from one axis and thresholds for face discrimination defined as the percent geometric variation required for subjects to perform at the 75% correct level. The increments were chosen to permit an accurate measurement of psychometric functions (Quick, 1974). As in previous experiments (Wilson et al., 2002), we did not find perceptual differences along different individual axes, so data were averaged. Means and standard errors of multiple runs (minimum of two performed on different days) are reported throughout. One of the authors

and two naïve subjects participated in this study. All had normal or corrected-to-normal vision.

2.2. Masks

To evaluate the nature of the masking, a variety of masks were tested. A noise mask (Fig. 1b) served as the baseline condition. It was created by applying the same band-pass filter used for the synthetic faces to a 2D binary noise array; the root-mean-square contrast was set to match that of the synthetic faces, which was 100% in all but one experiment. The mean synthetic face of the opposite gender, shown from a different viewpoint (20° to one side, see Fig. 1b) served as face mask. The size of the face was 50% larger than the test faces in order to minimise contour overlap between test and mask. Hence, when faces were masked by face stimuli, the stimuli always differed in gender, size and viewpoint.

Inverted and scrambled faces were employed to dissociate between local (feature based) and global (holistic) processes since both carry exactly the same total information as upright, intact faces but are perceived differently. Inverted face masks were identical to the upright masks up to a 180° rotation in the fronto-parallel plane (Fig. 3). Scrambled faces were generated by randomising the position of the features within a face leaving the head shape and hairline intact (Fig. 3). Additional masks were obtained by selectively removing either all internal features ('Head') or the head shape and hairline ('Features') from the face (Fig. 4). A band-pass filtered (same filter as that used for synthetic faces and noise mask) image of a grey-scale photograph of a house (Fig. 3) served as a mask of a non-face object category.

2.3. Apparatus

Stimuli were presented on an Apple iMac computer set to a spatial resolution of 1024×768 pixel and a frame rate of 75 Hz. The software lookup table was defined to maximise contrast linearity using 150 equally spaced grey levels. Pattern luminance was modulated about a mean of 38.0 cd/m^2. Subjects viewed the stimuli binocularly under dim room illumination, and a chin and forehead rest were used to maintain a constant viewing distance of 131 cm. At this distance each pixel subtended 0.012°. The program controlling the experiments included routines from the VideoToolbox (Pelli, 1997).

2.4. Procedure

The screen was set initially to a uniform grey field of mean luminance. Each trial was initiated by a mouse-click. The test face was flashed for a brief, 26.7 ms duration (2 frames), either following, or followed by, a mask of equal short duration with variable signal onset asynchrony (SOA). During the SOA period, the screen returned to the mean grey field. Following a 400 ms blank period, two faces were displayed simultaneously side by side and the subject indicated by a mouse-click on one of the faces which of the two matched the test (Fig. 1b). No time limit was placed on the decision process, but subjects rarely took more than 2.0 s. Feedback was not provided. SOAs were −240, −187, −133, −80, −54, 27, 80, 133, 187, and 240 ms.

3. Results

One aim of this study was to investigate the duration required for face processing. To do this, we measured the time course of the masking effect for two masks, faces and noise, on the discrimination of synthetic faces. The results are summarised in Fig. 2. Thresholds are the geometric information required to correctly discriminate two synthetic faces and are plotted relative to that for a noise mask at SOA = −240 ms. In a separate experiment we determined that the threshold obtained with a noise mask at SOA = −240 ms is no different from that obtained without mask. The dashed curve shows the performance when faces are masked by noise. It is clear from its flat appearance that noise has little or no masking effect on face discrimination, except for at 27 ms SOA (see later discussion on this point). In fact subjects

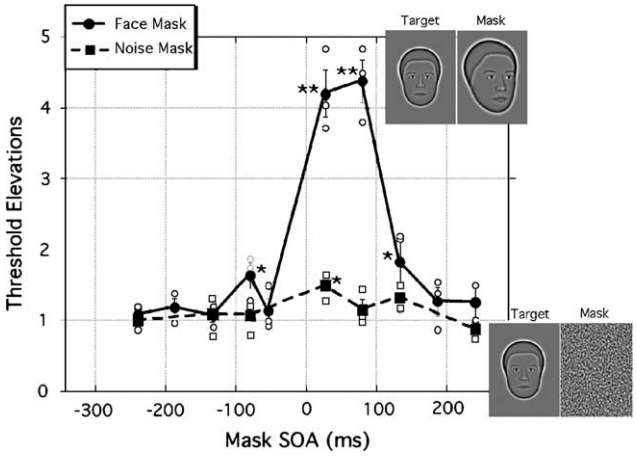

Fig. 2. Face discrimination in the presence of a noise (solid squares and dashed line) and a face (solid circles and line) mask as a function of mask SOA. Thresholds, measured as the geometric difference required to discriminate two faces, are plotted relative to that for a noise mask at SOA = −240 ms, which is the same as that obtained when no mask is used. Open symbols show data for individual subjects and filled symbols the average (squares for noise masks, circles for face masks). Here and elsewhere, bars represent standard errors of the mean. Insignificant masking effects for noise are in sharp contrast to face masks, which elevate thresholds up to a factor of 4.5 and exert a masking effect even when presented 133 ms after the target. Faces also show a small forward masking effect at SOA = −80 ms. Asterisks denote SOAs where discrimination was significantly impaired (*: $p < 0.05$; **: $p < 0.0001$).

frequently commented that they did not even consciously perceive the noise mask while focusing on the synthetic face. This is in sharp contrast to discrimination when the target face is presented with a face mask (solid curve). Thresholds depend strongly on SOA and are elevated up to 4.5 times for an SOA of 80 ms. Significant elevations are observed when a face mask is presented up to 133 ms after the onset of the target. There is also a small but significant effect when the masking face is presented prior to the target. This forward masking only occurs at an SOA of −80 ms. Here and elsewhere, differences were assessed statistically using a two-way, repeated measures ANOVA (subject by mask SOA). Differences were not significant across subjects ($F_{2,18} = 1.24$; $p = 0.3$) but highly significant for SOA ($F_{2,9} = 50.6$; $p < 0.0001$). Significant masking effects compared to the baseline were assessed by a post hoc analysis (Fisher's PLSD) and significant elevations are highlighted by asterisks in Fig. 2 (*: $p < 0.05$; **: $p < 0.0001$).

There are two important implications for the duration required for face discrimination. First, faces can be accurately discriminated for presentation times as short as 27 ms (performance for a 27 ms presentation is about as good as for 110 ms, Wilson et al., 2002) if they are unmasked or followed by noise. Second, and more importantly, face discrimination is impaired over a prolonged period if the target faces are masked by faces. The temporal window over which faces mask faces in our experiments suggests that it takes about 130 ms to complete the computation.

What is the cause of the dramatically different masking effects elicited by noise and faces? Since masking has often been attributed to low-level interactions between contours that are in close spatial and temporal proximity, we will consider contour interactions first. To compare the expected effect of noise and face masks, we calculated the amount of contour overlap between masks and targets in the following way. First, a typical face target and a mask (noise or face) were full-wave rectified relative to the mid-grey background, so as to capture any difference from the background, irrespective of it being above or below the mean luminance. The two stimuli (target and mask) were then superimposed and the fraction of pixels calculated at which both stimuli were different from the background. This calculation shows that a noise mask has about 2.3 times more contour overlap with the target than the face mask, which should not be surprising given that the face masks were always 50% bigger than the targets. Consequently, the observed differences in masking cannot be explained by the amount of contour overlap between target and mask.

A second possibility is that pattern similarity is responsible for the differences between noise and face masks. Our experimental design was chosen to minimise, as much as possible, the possibility of interference by a pattern (mask) that simply matched the target by using masks that differed from the targets in gender, size and orientation (see Fig. 1). This leaves as the most likely candidate object similarity, i.e. the fact that both target and mask are faces. We investigated this further in the remaining experiments.

The second part of our investigation was directed at elucidating the computations underlying face discrimination. One explanation for the difference in masking by noise and faces would involve interactions (or lack thereof) at a level where faces are encoded. If our results reflect interference at this level, little or no masking should be found when non-face objects are used as masks. Results using grey-scale images of houses as masks provide strong support for this hypothesis (Fig. 3). Performance is not measurably different from the noise mask.

How does masking depend on the arrangement of facial information? For example, scrambling the position of the internal features results in a stimulus that carries the same local feature information but alters the configural relationships among features. Similarly, inverted faces are physically similar but perceptually behave differently from upright faces (Diamond & Carey, 1986). If interruptions took place prior to face-specific processing (e.g. where contours are processed and contour interactions assumed to occur) we would expect to see an effect for scrambled and inverted faces similar to that for upright faces. Strikingly, both an inverted and a scrambled face exhibit significantly weaker masking than the upright face used in the first experiment. This emphasizes

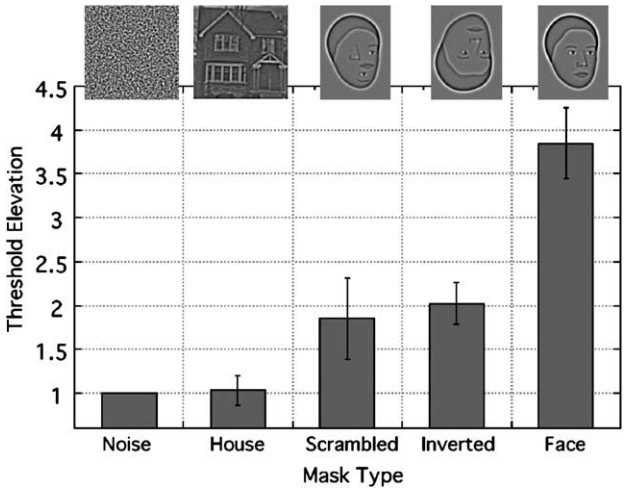

Fig. 3. Face discrimination in the presence of a variety of masks. Results are for an SOA of 80 ms where a face mask showed the strongest masking effect (see Fig. 2) and are relative to the performance of a noise mask. Objects from a different, non-face category (houses) show no masking. Both scrambled and inverted faces have a significant ($p < 0.01$) masking effect on synthetic face discrimination but the effect is much weaker than that observed with non-scrambled upright faces ($p < 0.0001$).

that configural similarity (“face-ness”) must be the major factor for this kind of masking and not the similarity of isolated facial features.

Based on these observations, the following question can be posed: what information contained within a face is critical for face masking? To investigate this, we compared the effect of three masks (Fig. 4): feature masks contained the internal features of the face without head shape or hairline; head masks had an intact head shape and hairline but no internal features; and full face masks. Three opposing predictions can be tested. If face parts were treated similar to full, intact faces, masking should be the same as for whole faces. Second, if they were processed differently (e.g. failing to activate face-specific processes), masking should be weak or absent. Finally, it is possible that either internal features alone or head shape alone might be the single cause of masking.

Results show the strongest masking effect for complete faces. Removal of either the head shape or the internal features from the face yields weaker masking but the masking effect in each case is still similarly stronger than a noise mask ($p < 0.05$). Although the features by themselves have only a small masking effect, their absence in the head shape mask significantly ($p < 0.0001$) reduced its masking effect. Thus, our data show that a full face generates significantly more masking than either features or head shape alone. Furthermore, the data indicate that the presence of features in the head shape produces a cooperative enhancement of masking.

The results so far highlight a masking effect that depends on configural similarity between target and mask. Masking is strong when similarity (face-ness) is preserved suggesting a locus of interaction where a specific object category (faces) is encoded. Noise presumably never reaches such a locus and hence does not impair processing. (Recall here that the noise mask was reported by subjects to be frequently invisible.) While noise may not cause excitation in regions specialised in face discriminations, it does, of course, evoke a response in early stages of visual processing. If our assumption about masking, as the interaction between a target and subsequent mask, is correct we should see evidence for interference at an early stage by noise, especially if a noise mask followed the target with a very short SOA. Indeed, Fig. 1 shows a small and just significant masking effect of noise at an SOA of 27 ms. The final experiment was designed to investigate further whether noise can have a significant masking effect on faces if potential interactions between mask and target are enhanced at an early stage. To do so, we reduced the contrast of the face targets to 10% (from previously 100%) while keeping the contrast of the noise at its maximum (100%). With this modification, noise has a highly significant masking effect on faces (Fig. 5). The magnitude of

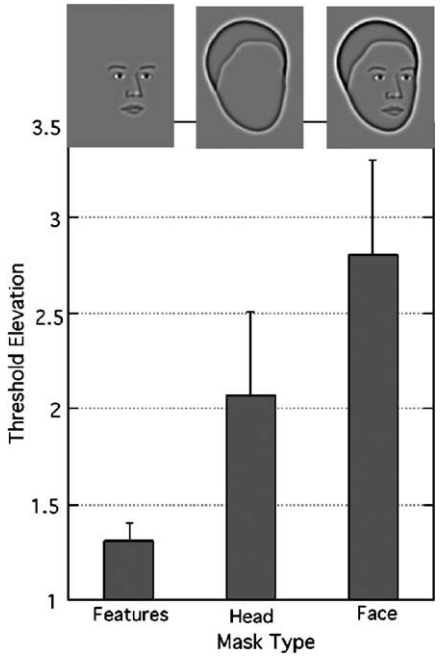

Fig. 4. Face discrimination when masks contain different details of a face (internal *features* without head shape; *head* shape and hairline without features; full *face*). Results are for an SOA of 80 ms and are relative to the performance with the noise mask. Data are averaged across experiments with front and side face views. Both, head shape and features alone yield more masking than noise ($p < 0.05$) but neither impairs performance to the extent observed with full faces ($p < 0.0001$). This suggests that the presence of features in the head shape produces a cooperative enhancement of masking.

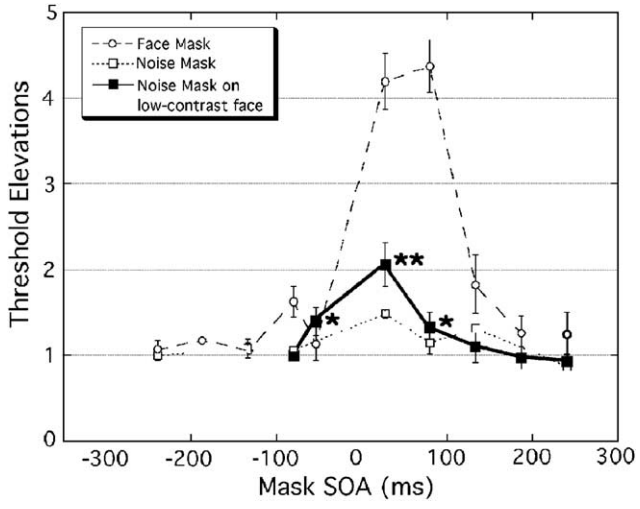

Fig. 5. The effect of high contrast noise on low contrast faces. A noise mask can mask faces if the contrast of the face targets is reduced (10%) while keeping the contrast of the noise high (100%). The bold solid curve shows that the masking effect is much weaker and extends over a much shorter period (about ±50 ms) than observed for face masks (dashed curve, re-plotted from Fig. 2 for comparison). Its effect is centred at the onset of the target. Also shown for comparison is the effect of a high contrast noise mask on a high contrast face target (dotted curve, re-plotted from Fig. 2). Asterisks indicate statistically significant masking effects for the high contrast mask on low contrast faces (*: $p < 0.05$; **: $p < 0.0001$).

this masking is, however, much lower than that observed for face masks (Fig. 5, dashed line), and its temporal window much shorter (⩽80 ms) peaking just after onset of the target face (27 ms). We hypothesize that this increased masking of low contrast faces by high contrast noise results from the increased latency when processing low contrast stimuli, thus permitting a slightly delayed, higher contrast mask to catch up with initial face contour processing at an early area (presumably V1).

4. Discussion

The aim of this study was to investigate the computations underlying face processing and their temporal dynamics using a masking paradigm. Central to our approach is the assumption that masking occurs if target and mask stimulate the same neural mechanisms and if the neural computation of the target is not completed when the response to the mask arrives. Performance is impaired as a consequence of interference between these two computations.

4.1. Duration of face processing

A straightforward application of masking is to determine the duration over which a mask has a detrimental effect on the stimulus and relate this to the timing of the underlying cortical computation. Once the target computation has been completed, any subsequently presented mask will leave perception unaffected. For face discrimination in our study, the duration over which a mask can impair performance is not a constant but depends on the type of mask.

When not masked by faces, synthetic face discrimination can be performed with a very short stimulus exposure. A 27 ms presentation time was sufficient in our experiments to reach peak performance; a longer (110 ms) presentation time, used in an earlier study (Wilson et al., 2002), shows no advantage. This duration is even shorter than one reported previously (between 50 and 100 ms) in a study on face discrimination, a difference that could be due to different stimuli (synthetic faces versus morphs of grey-scale face photographs, Lehky, 2000).

Regardless of these details, we concur that very short presentation times can be sufficient for face discrimination. This observation has lead to the proposal that faces are computed very rapidly (Lehky, 2000). Our results cast doubt on this proposal. It is evident when using face masks that a substantial time period is required before the computation is completed. Only when a face is followed by a non-face mask (noise in our case or a textured pattern in the earlier study), can the computation continue unmasked because the non-face pattern presumably never enters the locus where face processing occurs. However, this does not mean that face discrimination is completed rapidly. Our results suggest that face discrimination *appears* rapid when masked by stimuli that do not share sufficient similarity with the face targets. The prolonged duration required before face processing is completed can only be observed when face targets are masked by stimuli that are also faces. This duration in our study (about 130 ms) is in agreement with those reported for the recognition of facial emotions (Esteves & Ohman, 1993) and face identification (Costen et al., 1994). Both these studies masked their target faces with face masks (either with neutral expression or of different identity). Hence, it is conceivable that different aspects of face perception (face recognition, discrimination, perception of emotion) require a similar and substantial period for completion and that this time may be underestimated when face targets are masked by non-face stimuli.

Evidence for rapid face processing has also been found physiologically (Oram & Perrett, 1992). Neuronal responses were analyzed for their capability to predict the view of a face stimulus. Information theoretic analysis of spike trains in monkey STS shows that neurons can discriminate between face views differing by 45–60° within the first 5 ms of the response onset. While this is a remarkable achievement, it is a coarse task, which may not be comparable to face discrimination. Furthermore, view discrimination may be a necessary prerequisite for subsequent face recognition or discrimination. Discrimination may require additional computations, which take significantly longer. It would be interesting to see if discrimination between face views shows behaviourally shorter masking effects than those observed here.

In most pattern masking studies the detrimental effect of the mask peaks at short SOAs and extends over a very limited period (e.g. 40 ms, Kovacs et al., 1995), which is generally shorter than that observed for metacontrast masking (Enns, 2004). Such short-lifetime masking can be explained by contour interactions between target and mask in early visual areas. This cannot easily explain the much longer masking durations in our study. One possible explanation for our data is that mechanisms underlying face discrimination are slow and require time in excess of 100 ms. Alternatively, our 130 ms duration leaves enough time for recurrent computations in the form of feedback between, and/or lateral processing within, visual areas involved in face processing. This proposal is consistent with monkey physiology. Comparing earliest response latencies in striate cortex (35 ms in V1, Maunsell & Gibson, 1992) with those obtained in temporal cortex (60–69 ms in superior temporal sulcus, STS, Oram & Perrett, 1992), where units respond selectively to faces, shows that the signal can travel within about 25 ms between these areas in the ventral pathway (Lamme & Roelfsema, 2000).

This is short enough to allow for up to two recurrent iterations between V1 and STS within the prolonged masking duration in our experiments. Our results are moot on the exact nature of the competition (time consuming processes and/or feedback computations) underlying our masking effect. However, the extensive duration over which interactions occur challenges the hypothesis that face discrimination is completed rapidly within the fast feed-forward sweep of activation, which is assumed to reach the highest levels of the visual cortical processing hierarchy within 100 ms (Lamme & Roelfsema, 2000). Further evidence for the relatively slow processing of faces has been obtained from event related potential studies in which a major face-specific component (N170) occurs approximately 170 ms after stimulus presentation (Sagiv & Bentin, 2001).

4.2. Locus of masking and configural face processing

Two important questions concern the cortical site where masking occurs in our experiments and the type of computation carried out there. If masking in our experiments were due to interactions at an early visual stage (e.g. V1, Hubel & Wiesel, 1968) where contour orientations are processed, masking should be greatest for noise and about the same for a full face and a scrambled or inverted mask, since noise has the highest amount of contour overlap. This is clearly not supported by our results. If masking were due to interactions at a (hypothetical) intermediate level where individual face parts were processed (e.g. head contour as a global shape possibly in area V4, Gallant, Connor, Rakshit, Lewis, & VanEssen, 1996; Wilkinson et al., 2000), scrambled and inverted masks should yield similar masking to a full face. This is also not supported by our data. Instead, strong masking only occurs when the mask matches the face globally (features placed correctly within a head). This is the hallmark of holistic processing (Diamond & Carey, 1986; Tanaka & Farah, 1993) and shows that the strong masking depends on the configural similarity between target and mask. This leaves as an obvious location of interactions a stage where faces are encoded as a whole (possibly FFA).

Two alternative explanations must be considered. First, since masking correlates with the degree of similarity between target and mask, it seems tempting to argue that our masking has little to do with face processing per se but is instead due to interruptions when attempting to match (or contrast) two similar patterns. Simple pattern matching effects as the main reason for masking are unlikely because the face masks were always bigger than the target face, shown as a side view rather than a front-view, and were from the opposite gender. Given this, there is little pattern similarity between target and face mask. Also, there is little difference in pattern similarity between the targets and the upright versus inverted faces. This adds further weight to the assumption that our masking happens at a level of face processing where neurons respond to faces of different gender, different size and viewpoint.

Second, it is possible that different masks interfere at different levels of processing. For example, the reduced but significant masking with scrambled and inverted faces could either be due to interference at a higher level where face-specific computation takes place or at an intermediate stage (e.g. where isolated face parts are first extracted from an image and subsequently form the input to holistic processes). In the former scenario, scrambled and inverted faces would disrupt face computation but to a lesser extent than full faces. In the latter, early interference may result in weaker overall disruption. Our results are moot on the exact location where scrambled and inverted faces mask whole upright faces but, according to imaging and physiological studies, it seems clear that inverted faces do elicit signals, albeit weaker than whole upright faces, in areas specialised in face processing. FMRI investigations on FFA (Kanwisher, Tong, & Nakayama, 1998), face-specific ERPs (Jeffreys, 1989) and recordings from face selective neurons in monkey STS (Perrett et al., 1988) reported differences, often small, between response magnitude and latencies of upright and inverted faces. Behavioural results have also provided evidence that inverted and upright faces engage the same mechanism (Sekuler, Gaspar, Gold, & Bennett, 2004). Using an image classification technique, Sekuler et al. found no qualitative difference in the kind of information used to discriminate upright or inverted faces. The information was simply used less efficiently in inverted faces, yielding the typical inversion effect with higher sensitivity to upright faces. If the same mechanisms are excited by inverted and upright faces but the information from inverted faces is combined less efficiently, the neuronal response for inverted faces would be expected to be weaker than that for upright faces. This would explain why we see a significant masking effect by inverted faces and why this effect is weaker than that observed with upright faces. Such a mechanism might give more masking when inverted faces are masked by upright faces than by inverted faces, a prediction that could be tested experimentally. The evidence from these studies makes it likely that inverted and scrambled masks disrupt processing at a stage where faces are encoded and that these interactions give rise to the small but significant masking in our experiments.

Reduced masking for scrambled faces relative to intact faces supports the existence of holistic processes, a conclusion that was also reached by Farah and colleagues (Farah, Wilson, Drain, & Tanaka, 1998). In a face matching experiment, whole faces elicited stronger masking than scrambled face masks. Interestingly, the difference between scrambled and inverted masks was much less pronounced for other objects classes such as

houses or words (when masked by words versus scrambled letters), indicating that holistic representations are unique (or particularly important) for upright faces (Farah et al., 1998; Johnston & McClelland, 1980).

We investigated the nature of holistic computations further by measuring the masking effect observed when reducing the amount of information contained in an upright face mask. If face features were combined into a holistic representation with no part-based representation, isolated face parts should give little or no masking. This would also be the case if face features were processed independently from holistic faces, a view that has received support from behavioural (Hillger & Koenig, 1991), neuropsychological (Yin, 1970) and brain imaging studies (Rossion et al., 1999). Our results argue against these strategies because face parts impair face processing. The level of masking is presumably related to the degree to which isolated features activate (and therefore interfere with) face discrimination mechanisms. Both internal features and head shape mask but their effect is weaker than that for the full face. These masking data on isolated face parts are in qualitative agreement with monkey physiology (Desimone, Albright, Gross, & Bruce, 1984). Face selective neurons in IT cortex respond best to full faces but also (albeit less) when eyes or noses are eliminated.

It is conceivable, as suggested by one reviewer, that these results on face parts are not specific to face processing but rather a more general feature of object recognition. Consistent with this idea, a complete house would mask more, in a house discrimination task, than a house without windows and doors. While this is a possibility, we believe there are several factors that argue against this interpretation. Firstly, the fact that inverted and scrambled faces mask less than upright and unscrambled faces suggests that our masking effect occurs at a level where faces are processed, since inversion effects have been shown to be particularly strong for faces compared to other object classes. Secondly, there is experimental evidence showing that scrambling the 'internal' features of house masks (i.e. location of windows and doors) does not strongly diminish house recognition compared to intact house masks (Farah et al., 1998). This is in contrast to faces in our study and earlier investigations (Farah et al., 1998) where scrambling reduces the masking effect. This suggests that our results are likely to be due to interactions within face-specific computations rather than a general feature of object categorisation.

The reduction in masking whenever the masking stimulus is not an upright, whole face (inverted, scrambled, isolated face parts) supports the notion of holistic processing. These results, however, do not support exclusively holistic processing, that responded only to the simultaneous presence of features and head at the appropriate locations. Because all modified face stimuli yield masking beyond that of noise or house masks, they must interrupt whole face processing suggesting that they, unlike non-face stimuli, engage similar neural mechanisms to upright intact faces.

4.3. Theories for masking

Various models for the detrimental effect of backward masking have been proposed (Breitmeyer & Ganz, 1976; Di Lollo, 1980; Turvey, 1973). Motivated by shortcomings of all these models, a modified theory was presented recently (Di Lollo, Enns, & Rensink, 2000; Enns, 2004), including two distinct components of masking. One component is early or fast acting ($<$100 ms) associated with object formation, the second later or slow acting ($\geqslant$150 ms) associated with object substitution. The later component was thought to occur because one object is replaced with another in conscious representation but distributed attention (as a function of the number of distracters in which the target was embedded) was required for object substitution to be observed. If object substitution is a universal quantity, we presumably did not observe it because there was no spatial uncertainty about the target and hence no divided attention in our experiments. Moreover, our masking effect does not fall into the temporal bracket for object substitution.

One of the problems posed by our experiments for standard theories is the fact that object similarity is a critical parameter and determines the magnitude of masking. Similarity effects have also been observed in other studies. Enns (2004) showed that letter identification is most impaired when masked by letters. Both digits and noise also masked but to a lesser extent; four dots surrounding the target location showed no masking. In contrast to our results, the duration over which masking occurred in Enns' study was the same for different masks and taken as evidence that all masking was due to the same mechanism. Differences in magnitudes were interpreted to result from low-level temporal integration of information arising from the target and the mask. According to this view, the strength of masking depends on the number of target relevant features contained in the mask. It is interesting to note that masking does not depend on target-mask similarity in word recognition. Different masks, either consisting of whole words or made up of scrambled letters, yield about the same masking effect (Farah et al., 1998). We believe that the similarity effect in face discrimination in our experiments is due to the configural nature of face coding, a strategy that is employed for faces but not for other object categories.

What neural mechanisms could underlie these masking results? Based on electrophysiological observations, the effect of a mask, presented very shortly after a target, is given by a reduced firing rate and an altered tuning

profile (Kovacs et al., 1995; Rolls & Tovee, 1994; Rolls, Tovee, & Panzeri, 1999). This has been explained by the assumption that interference occurs as long as the mask reaches the neural site within the initial response transient of the target. This could account for the masking effect of noise at an early level, where both target and mask trigger responses from filters tuned to edge orientation (DeValois, Yund, & Hepler, 1982; Hubel & Wiesel, 1968). This effect should be particularly evident if a low contrast target is followed by a high contrast mask, assuming that low contrast stimuli require longer processing time, and this is exactly what we have observed. Masking at this early stage could be due to interactions at the level of contrast gain control mechanisms, which are assumed to be driven by the concerted activity of neurons, largely independent of preferred orientation, spatial frequency and location (Heeger, 1992). Similarly, face-like stimuli (upright, inverted, scrambled or isolated face parts) could mask faces by analogous competitive neural interactions at a higher level where faces are encoded.

Our results therefore suggest a modification to the model proposed by Enns (2004). The component associated with object formation should be extended to include more than one level of computation, a notion consistent with the hierarchical architecture of the visual system. Our masking paradigm offers a way to experimentally investigate different levels of visual processing.

5. Conclusions

In conclusion, our results show a strong effect of configural similarity. Upright faces, even when shown from a different viewpoint, with different size and gender, exert a dramatic backward masking effect on face discrimination. Isolated face parts as well as inverted and scrambled faces show significantly weaker masking. Objects from a different category (houses) exhibit no masking under these circumstances. These data can only be explained by configural pattern similarity and not by degree of contour overlap, consistent with an extra-striate masking locus, possibly FFA. The duration over which masking occurs is substantial, suggesting that face discrimination is a time consuming process, which may not be completed rapidly in the fast feed-forward sweep of activation. Our results demonstrate that configural masking provides a powerful psychophysical tool for selectively investigating different processing stages in the study of higher-level form vision.

Acknowledgments

We are indebted to A. Gorbatski for serving as subject and aiding with data analysis and two anonymous referees whose comments helped substantially to improve this manuscript. This research was supported in part by EPSRC grants #GR/S62666/01 and #GR/S59239/01 to GL, NIH grant #EY002158 to HRW and NSERC grant #OP0007551 to FW.

References

Allison, T., Puce, A., Spencer, D. D., & McCarthy, G. (1999). Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cerebral Cortex, 9*, 415–430.

Breitmeyer, B. (1984). *Visual masking: An integrative approach*. New York: Oxford University Press.

Breitmeyer, B. G., & Ganz, L. (1976). Implications of sustained and transient channels for theories of visual-pattern masking, saccadic suppression, and information-processing. *Psychological Review, 83*, 1–36.

Bruce, V., & Young, A. (1998). *In the eye of the beholder: The science of face perception*. Oxford: Oxford University Press.

Collishaw, S. M., & Hole, G. J. (2000). Featural and configurational processes in the recognition of faces of different familiarity. *Perception, 29*, 893–909.

Costen, N. P., Shepherd, J. W., Ellis, H. D., & Craw, I. (1994). Masking of faces by facial and non-facial stimuli. *Visual Cognition, 1*, 227–251.

Desimone, R. (1991). Face-selective cells in the temporal cortex of monkeys. *Journal of Cognitive Neuroscience, 3*, 1–8.

Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience, 4*, 2051–2062.

DeValois, R. L., Yund, E. W., & Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research, 22*, 531–544.

Diamond, R., & Carey, S. (1986). Why faces are and are not special—an effect of expertise. *Journal of Experimental Psychology—General, 115*, 107–117.

Di Lollo, V. (1980). Temporal integration in visual memory. *Journal of Experimental Psychology—General, 109*, 75–97.

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology—General, 129*, 481–507.

Enns, J. T. (2004). Object substitution and its relation to other forms of visual masking. *Vision Research, 44*, 1321–1331.

Esteves, F., & Ohman, A. (1993). Masking the face—recognition of emotional facial expressions as a function of the parameters of backward-masking. *Scandinavian Journal of Psychology, 34*, 1–18.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception?. *Psychological Review, 105*, 482–498.

Gallant, J. L., Connor, C. E., Rakshit, S., Lewis, J. W., & VanEssen, D. C. (1996). Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *Journal of Neurophysiology, 76*, 2718–2739.

Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Research, 39*, 3537–3560.

Gross, C. G. (1992). Representation of visual-stimuli in inferior temporal cortex. *Philosophical Transactions of the Royal Society of London Series B—Biological Sciences, 335*, 3–10.

Gross, C. G., Rocha-Miranda, C. E., & Bener, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology, 35*, 96–111.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience, 9*, 181–197.

Hillger, L. A., & Koenig, O. (1991). Separable mechanisms in face processing—evidence from hemispheric-specialization. *Journal of Cognitive Neuroscience, 3*, 42–58.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of the monkey striate cortex. *Journal of Physiology, 195*, 215–243.

Jeffreys, D. A. (1989). A face-responsive potential recorded from the human scalp. *Experimental Brain Research, 78*, 193–202.

Johnston, J. C., & McClelland, J. C. (1980). Experimental tests of a hierarchical model of word identification. *Journal of Verbal Learning and Verbal Behavior, 19*, 503–524.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*, 4302–4311.

Kanwisher, N., Tong, F., & Nakayama, K. (1998). The effect of face inversion on the human fusiform face area. *Cognition, 68*, B1–B11.

Kovacs, G., Vogels, R., & Orban, G. A. (1995). Cortical correlate of pattern backward-masking. *Proceedings of the National Academy of Sciences of the United States of America, 92*, 5587–5591.

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23*, 571–579.

Lehky, S. R. (2000). Fine discrimination of faces can be performed rapidly. *Journal of Cognitive Neuroscience, 12*, 848–855.

Loffler, G., Wilkinson, F., Yourganov, G., & Wilson, H. R. (2004). Effect of facial geometry on the fMRI signal in the fusiform face area [abstract]. *Journal of Vision, 4*(8), 136a. Available from http://journalofvision.org/4/8/136/.

Maunsell, J. H. R., & Gibson, J. R. (1992). Visual response latencies in striate cortex of the macaque monkey. *Journal of Neurophysiology, 68*, 1332–1344.

Moscovitch, M., & Radzins, M. (1987). Backward-masking of lateralized faces by noise, pattern, and spatial-frequency. *Brain and Cognition, 6*, 72–90.

Nasanen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision Research, 39*, 3824–3833.

Oram, M. W., & Perrett, D. I. (1992). Time course of neural responses discriminating different views of the face and head. *Journal of Neurophysiology, 68*, 70–84.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Perrett, D. I., Mistlin, A. J., Chitty, A. J., Smith, P. A. J., Potter, D. D., Broennimann, R., et al. (1988). Specialized face processing and hemispheric-asymmetry in man and monkey—evidence from single unit and reaction-time studies. *Behavioural Brain Research, 29*, 245–258.

Quick, R. F. (1974). A vector-magnitude model of contrast detection. *Kybernetik, 16*, 65–67.

Regan, D. (2000). *Human perception of objects*. Sunderland, MA: Sinauer.

Rolls, E. T., & Tovee, M. J. (1994). Processing speed in the cerebral-cortex and the neurophysiology of visual masking. *Proceedings of the Royal Society of London Series B—Biological Sciences, 257*, 9–15.

Rolls, E. T., Tovee, M. J., & Panzeri, S. (1999). The neurophysiology of backward visual masking: Information analysis. *Journal of Cognitive Neuroscience, 11*, 300–311.

Rossion, B., Delvenne, J., Debatisse, D., Goffaux, V., Bruyer, R., Crommelinck, M., et al. (1999). Spatio-temporal localization of the face inversion effect: An event-related potentials study. *Brain, 115*, 15–36.

Sagiv, N., & Bentin, S. (2001). Structural encoding of human and schematic faces: Holistic and part-based processes. *Journal of Cognitive Neuroscience, 13*, 937–951.

Sekuler, A. B., Gaspar, C. M., Gold, J. M., & Bennett, P. J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology, 14*, 391–396.

Sergent, J., & Signoret, J. L. (1992). Functional and anatomical decomposition of face processing—evidence from prosopagnosia and pet study of normal subjects. *Philosophical Transactions of the Royal Society of London Series B—Biological Sciences, 335*, 55–62.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology Section A—Human Experimental Psychology, 46*, 225–245.

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition, 25*, 583–592.

Turvey, M. T. (1973). On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review, 81*, 1–52.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology Section A—Human Experimental Psychology, 43*, 161–204.

Wilkinson, F., James, T. W., Wilson, H. R., Gati, J. S., Menon, R. S., & Goodale, M. A. (2000). An fMRI study of the selective activation of human extrastriate form vision areas by radial and concentric gratings. *Current Biology, 10*, 1455–1458.

Wilson, H. R., Loffler, G., & Wilkinson, F. (2002). Synthetic faces, face cubes, and the geometry of face space. *Vision Research, 42*, 2909–2923.

Yin, R. K. (1970). Face recognition by brain-injured patients: A dissociable ability. *Neuropsychologia, 8*, 395–402.